

**Northeast Region I&M Program
QA/QC Procedures and Metadata Development for Final Biological
Inventory Products**

DRAFT

September 8, 2004

INTRODUCTION

The inventory and monitoring projects sponsored by the Northeast Coastal and Barrier Network produce high-quality biological and spatial datasets to support effective management of the network's natural and cultural resources.

This document outlines the procedure used to review these datasets. A Principal Investigator at a local university or agency often creates these datasets. The detailed review of biological inventory data is important to ensure the quality of data submitted to the National Park Service.

In short, the review includes checking the spatial data, ensuring that spatial metadata exists, and creating biological metadata for the tabular biological inventory data. Metadata is important because it documents the creation, processing, and storage of the datasets.

DATA FLOW

The Principal Investigator submits the data to the appropriate Network Data Manager.

The Network Data Manager does a preliminary review of the data. When the following items have been collected the Data Manager submits all items as one zip file to the North Carolina State University (NCSU) for a full review and the creation of biological metadata.

1. Final Report and all relevant tables and appendices
2. All relevant spatial data (shapefiles) including metadata when available
3. All raw data which will be in Microsoft Excel or Access
4. A data dictionary for the raw data when available

NCSU is responsible for submitting the reviewed data files to the Data Manager on a CD with FGDC compliant biological metadata.

DATA ORGANIZATION

All biological inventory data that is received by NCSU is stored in a folder with the same name as the original zip file name. The zip file name is assigned by the I&M Data Manager and includes a unique combination of the cooperators name or organization, the park or network, the taxon, and the date. An example is vins_netn_122203.zip.

This folder is stored in three directories:

1. backup_original_inventories
2. working_inventories
3. completed_inventories

The backup_original_inventories directory houses all the original data submitted to NCSU. It is not organized.

The working_inventories directory houses all the working data and is organized into four subfolders:

final_report	The final report for the project including all tables and appendices.
gis_data	All shapefiles and/or other spatial data submitted.
raw_data	All raw data including Microsoft Access database, Microsoft Excel spreadsheets, and any other relevant raw data (for example, recorded bat calls).
Metadata	Both biological and spatial metadata records.

When the biological inventory review and metadata are completed, the folder and four subfolders are moved to the completed inventories directory.

A copy of the folder moved to the completed inventories directory is burned on to a CD and submitted to the data manager.

FINAL REPORT REVIEW

I&M Network Staff is responsible for reviewing content and format of all final reports.

The NCSU review includes:

1. Opening the file and verifying that the final report is not a draft report.

2. Comparing the Table of Contents to the opened report to ensure that all tables, appendices, and maps are included. These may or may not be separate files.

SPATIAL DATA REVIEW

The spatial data for the biological inventories consists of survey locations. These can be points, lines, or polygons. All spatial data should be projected in the local UTM zone using the NAD83 datum. If shapefiles are submitted, they should be organized by park.

VERIFY X,Y COORDINATES IN LOCATIONS TABLE

The x,y coordinates for each survey location are stored in the main locations table (tblLocations). Verify that these coordinates are correct by using them to create a shapefile. This should be done in ArcMap using Tools | Add XY Data. If a table has x,y coordinates for more than one UTM zone, create a shapefile for each zone.

Compare the new shapefile(s) to the shapefile(s) submitted by the principal investigator. Verify the number of locations in each shapefile and that the locations overlap.

CREATE SHAPEFILES

If shapefiles were not submitted by the principal investigator, use the shapefile(s) created above. Create a separate shapefile for each park.

REVIEW SHAPEFILES

The final shapefile(s) should be reviewed using the following questions:

1. Do all features line up with other park data such as the park boundary and the orthophoto mosaic?
2. Do all features fall within the park boundary?
3. Are all polygons complete? Any slivers?
4. Is the shapefile in the correct UTM zone and datum?
5. Do the survey locations match the survey locations in the raw data? Compare the number of survey locations in each dataset.
6. Do the number of survey locations in the shapefile match the number of survey locations in the final report?

7. Does the attribute table have a key field to link to the raw data like LocationID? If not, this needs to be added.
8. Is the data type of the key field in the attribute table the same data type as the key field in the raw data? This is important for linking the raw data to the spatial layers using GIS.
9. Do any fields in the attribute tables have unacceptable characters i.e., characters other than letters, numbers, and underscores? Make sure there are no spaces in the field or file names!
10. Are data values truncated because field lengths are too short?
11. Does each field have only one value?

SPATIAL METADATA

If spatial metadata is submitted for the shapefiles, review the metadata and make sure that it is FGDC-compliant as outlined in the Content Standard for Digital Geospatial Metadata (FGDC-STD-001-1998).

If the spatial metadata is not submitted, contact the cooperator and ask them to fill-out the Northeast Region GPS Metadata Form (Appendix XXX). Use the information from the completed form to create FGDC-compliant metadata using the Metadata Editor in ArcCatalog.

After exporting the metadata as a text file from ArcCatalog:

1. Delete the Online Linkage section
2. Add shapefile name under Resource Description in the Distribution section
3. Delete everything after Standard Order Process in the same section
4. Delete import metadata process step
5. Delete metadata extension information

Parse all metadata records using the latest version of the Metadata Parser (mp).
Correct any errors.

RAW DATA REVIEW

1. Print and review the final report and note the number of parks and each type of survey included.
2. If raw data is not in Microsoft Access, import it into Access. The purpose of importing the raw data into Access is to create a look-up table for the species. The species table is used to create a text file of species names that is then uploaded to the online ITIS Taxonomic Metadata Tool. The output is downloaded and inserted

in the taxonomy section of the biological metadata. Converting the data into Access also ensures that the data does not have any special formatting (bold text, summary columns), as is often the case in Microsoft Excel. This facilitates documenting the tables in the entity and attribute section of the biological metadata as well as linking the data in a GIS.

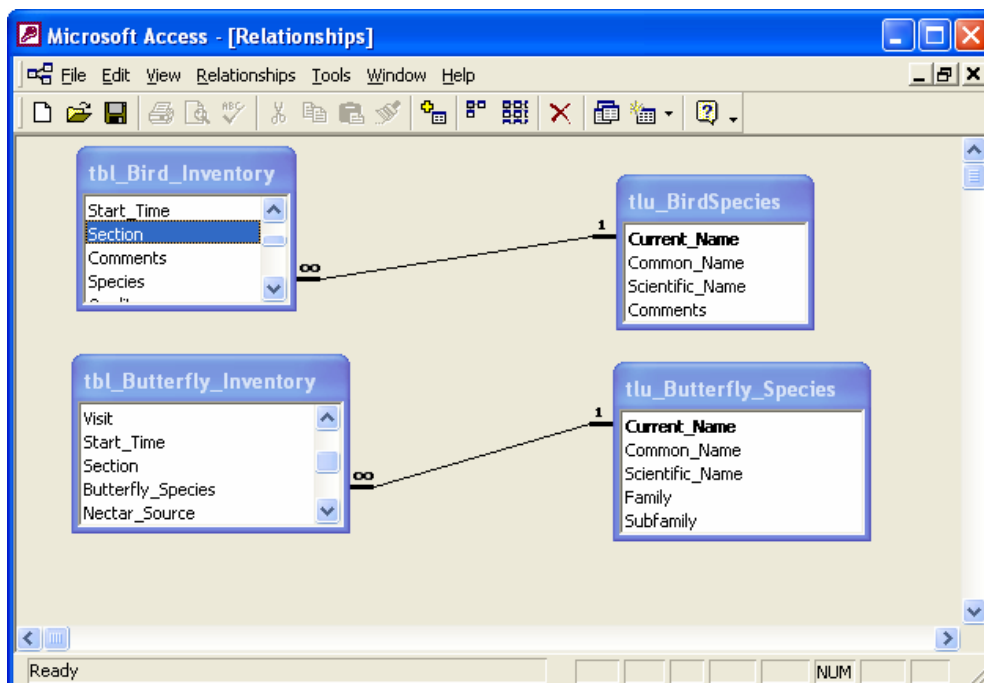
3. Review the final raw data in Access.
 - a. Are all surveys included?
 - b. Are all parks included for each survey?
 - c. Do the numbers of survey locations for each survey/park match the numbers in the final report?
4. Compare methodology descriptions to tables in the database to make sure they make sense.
5. At the minimum, create tables in Access for survey locations, for each survey, and a look-up table for each taxon.
6. Make sure the field and table names are appropriate. Do not use characters or spaces. Underscores are acceptable.
7. Use the raw data and final report to create look-up tables. The look-up tables should include species. These will be used to complete the taxonomy section of the biological metadata.

An example species look-up table for species is:

Field Name	Field Definition
Current_Name	This is the name used by the investigator to document the species in the data table. This can be a four-letter code, a common name, or a scientific name.
Common_Name	This is the accepted common name for the species as documented in the final report.
Scientific_Name	This is the scientific name documented by the investigator in the final report.
ITIS_Name	This is the scientific name recognized by ITIS. This may be the same as the scientific name in the final report. Discrepancies are documented in the taxonomic section of the biological metadata.

Other fields can include family, subfamily, comments, or any other fields that will make filling out the taxonomy section easier.

8. Create relationships and enforce referential integrity between the survey tables and the species look-up tables to make sure you have not missed any species.



9. Use information from the final report to complete the database descriptions in Access if a data dictionary is not included. Enter descriptions for each field for each table. Gather the relevant information from the final report. If an adequate description is not available, contact the principal investigator. This data will be used to complete the entity and attribute information section of the biological metadata.

BIOLOGICAL METADATA

Biological metadata is used to document the tabular data for each biological inventory. The standard is documented in the "Content Standard for Digital Geospatial Metadata, Part 1: Biological Data Profile" (FGDC-STD-001.1-1999).

A detailed procedure log for creating biological metadata for the biological inventory data is outlined in the stand-alone guidance document title "Northeast Region I&M Program Standard Operating Procedure: For Metadata Tools Use in the Creation of the Biological Data Profile of the Content Standard for Digital Geospatial Metadata." Appendix XXX.

COMMUNICATION

Deleted: ¶

In the event that information is missing, the appropriate Network Data Manager is contacted by NCSU with a list of items. Either the Data Manager or staff at NCSU contacts the biological inventory Principal Investigator for additional information.

VERSION CONTROL

If NCSU receives additional items for a biological inventory after the initial receipt of the original zip file, the additional items will be archived in a dated folder in the backup_original_inventories folder before being added to the working_inventories folder.

Naming conventions for versioned items will include the name of the file followed by the date in this format "..._20040826."